The Role of Computerized Diagnostic Proposals in the Interpretation of the 12-lead Electrocardiogram by Cardiology and Non-Cardiology Fellows

The Role of Computerized Diagnostic Proposals in the Interpretation of the 12-lead Electrocardiogram by Cardiology and Non-Cardiology Fellows

Tomas Novotny, MD, PhD,[a] Raymond Bond, PhD,[b] Irena Andrsova, MD, PhD,[a]
Lumir Koc, MD,[a] Martina Sisakova, MD,PhD,[a]
Dewar Finlay, PhD,[b] Daniel Guldenring, PhD,[b]
Jindrich Spinar, MD, PhD,[a] and Marek Malik, PhD, MD[c]

[a] Department of Internal Medicine and Cardiology, University Hospital Brno and Faculty of Medicine of Masaryk University, Brno, Czech Republic
[b] Faculty of Computing and Engineering, Ulster University, United Kingdom
[c] St. Paul's Cardiac Electrophysiology and Imperial College, London, United Kingdom

Address for correspondence:
Tomas Novotny, MD, PhD
Department of Internal Medicine and Cardiology
University Hospital Brno
Jihlavska 20
62500 Brno
Czech Republic
Email: novotny-t@seznam.cz

Highlights

- based on a total of 9000 ECG interpretations it was shown that computerized diagnostic proposals affect both the diagnostic accuracy and the interpreters´ confidence in their conclusive diagnosis
- the accuracy is significantly influenced especially when a single diagnostic proposal (either correct or incorrect) is provided – - giving the correct diagnosis improves the accuracy while giving a wrong diagnosis lowers the accuracy
- presentation of multiple computerized diagnoses improved the diagnostic accuracy of ECG interpreters

ABSTRACT

Introduction: Most contemporary 12-lead electrocardiogram (ECG) devices offer computerized diagnostic proposals. The reliability of these automated diagnoses is limited. It has been suggested that incorrect computer advice can influence physician decision-making. This study analyzed the role of diagnostic proposals in the decision process by a group of fellows of cardiology and other internal medicine subspecialties.

Materials and methods: A set of 100 clinical 12-lead ECG tracings was selected covering both normal cases and common abnormalities. A team of 15 junior Cardiology Fellows and 15 Non-Cardiology Fellows interpreted the ECGs in 3 phases: without any diagnostic proposal, with a single diagnostic proposal (half of them intentionally incorrect), and with four diagnostic proposals (only one of them being correct) for each ECG. Self-rated confidence of each interpretation was collected.

Results: Availability of diagnostic proposals significantly increased the diagnostic accuracy ($p<0.001$). Nevertheless, in case of a single proposal (either correct or incorrect) the increase of accuracy was present in interpretations with correct diagnostic proposals, while the accuracy was substantially reduced with incorrect proposals. Confidence levels poorly correlated with interpretation scores ($rho\approx 2$, $p<0.001$). Logistic regression showed that an interpreter is most likely to be correct when the ECG offers a correct diagnostic proposal (OR=10.87) or multiple proposals (OR=4.43).

Conclusion: Diagnostic proposals affect the diagnostic accuracy of ECG interpretations. The accuracy is significantly influenced especially when a single diagnostic proposal (either

correct or incorrect) is provided. The study suggests that the presentation of multiple computerized diagnoses is likely to improve the diagnostic accuracy of interpreters.

Keywords: computerized diagnostic proposals; decision making; electrocardiogram interpretations;

1. INTRODUCTION

Practically all contemporary 12-lead electrocardiogram (ECG) devices offer automatic computerized diagnostic proposals to assist diagnostic decision-making. However, the reliability of these automated diagnostic proposals is still sub-optimal.[1] Contrary to many other fields, the human interpretation of ECG tracings (by a cardiologist) still outperforms the diagnostic accuracy of computer diagnostic programs and artificial intelligence systems.[2] Computer programs frequently misdiagnose ECGs showing pathological cardiac rhythms.[3,4] As a result, the recently published recommendations concluded that all computer-based ECG interpretations require physician over-reading.[3] Several smaller studies have suggested that incorrect computer advice can influence physician decision-making and can lead to additional unnecessary diagnostic testing and/or inappropriate therapy.[4,5] Having this in mind, we aimed at analyzing the role of the provision of diagnostic proposals in the diagnostic decision-making process by a group of junior fellows trained in cardiology and another group of fellows trained in other internal medicine subspecialties. The study was a continuation of our previous investigation.[6]

2. METHODS

2.1. ECGs and study participants

A set of 100 clinical 12-lead ECG tracings were selected and grouped into 12 diagnostic classes (Table 1) covering both normal cases and common abnormalities. The true diagnostic meaning of the selected ECGs was based on the consensus of three experienced cardiologists. Of the 100 tracings, 23 were classified as representing life threatening conditions of acute coronary syndrome and hemodynamically compromising arrhythmias. All ECG tracings were printed on paper in the standard format of clinical ECGs (using printout layouts produced by different ECG equipment – that is 4 columns of 3 ECG leads, 2 columns of 6 leads, and 1

column of 12 leads) and their computerized diagnoses and interval measurements were removed in order to control the experiment.

A team of 15 junior Cardiology Fellows (CFs) and 15 Non-Cardiology Fellows (non-CFs) were recruited without prior assessment of their electrocardiographic knowledge (Table 2). On average, these fellows were in their third year of post-graduate training. All fellows have regular duties within the general emergency department at the University Hospital Brno, where they regularly encounter a wide range of clinical situations requiring ECG interpretation. Prior to the study, the fellows were trained in electrocardiology during regular consultations and by self-study. No additional training above this level was provided.

The ECG tracings were distributed to the fellows in randomly selected batches of 20 good quality paper copies. The fellows were asked to diagnose each ECG without any consultation and to give a self-rated confidence level to indicate their belief that their interpretation was correct (where 1=not very confident, 10=very confident). In total, each fellow had to interpret the same 100 ECGs three times in each of the three phases. In Phase 1, the ECGs were provided with no diagnostic proposals and the fellows had to submit their diagnoses in writing. In the Phase 2, one diagnostic proposal was provided for each ECG tracing with one half of the proposals being intentionally incorrect. The fellows were asked to either confirm the proposal or dismiss it and write their own diagnosis. In the Phase 3, four different diagnostic proposals were provided for each ECG tracing only one of which was correct. The fellows were asked to select the correct proposal. The fellows were diagnosing the ECGs without any consultation and/or reference searches. No time limits were imposed for an interpretation of the ECG tracing and no time measurements were taken.

The phases were separated by four-week intervals. To remove a confounding bias, counter-balancing was used by one half of the participants starting Phase 1 followed by Phase 2, whilst the other half started with Phase 2 followed by Phase 1. In both cases, Phase 3 was conducted as the last phase of the experiment. The participants did not receive any feedback about their results until the end of the entire study.

2.2. Data analysis

Diagnostic interpretations by all fellows were reviewed by two experienced cardiologists and scored as: (1) correct, (2) almost correct (i.e. the essential diagnosis made correctly with minor ECG details omitted), (3) incorrect, and (4) dangerously incorrect (i.e. seriously incorrect classification - either a wrong diagnosis that would not lead to proper treatment in cases where an immediate treatment or interventions were needed, or a wrong diagnosis that would lead to unnecessary treatment in cases when a non-existent pathology was diagnosed likely implying unnecessary and/or possibly dangerous treatment or measures.) In other words, diagnostic errors that might likely lead to severe clinical consequences were termed "dangerously incorrect". By this term, we merely imply that the misdiagnoses might potentially have dangerous consequences in some but not necessarily all clinical cases. Examples of what we considered a dangerously incorrect interpretation are presented in Figure 1. All participants of the study provided their interpretations in Czech and the classifications of the diagnoses were made by Czech experienced cardiologists, excluding any possibility of language and/or translation-related errors.

Diagnostic accuracies and other proportions are presented as percentages. The Mann-Whitney test was used to compare the performance of the CF and non-CF cohorts, the Wilcoxon test was used for paired comparisons. Multivariate logistic regression was used to identify statistically significant variables (using Chi-square testing) that increase or decrease the likelihood of a correct ECG interpretation (as determined by the odds ratios). This was also used to identify any confounding variables that might affect the results. Spearman correlation coefficient was used to measure the association between interpretation score and confidence ratings. Confidence intervals were derived and presented where necessary. All data analysis was performed using the R programming language and R Studio. P-values below 0.05 were considered statistically significant.

3. RESULTS

3.1. Diagnostic accuracies

A total of 9000 ECG interpretations were collected. Figure 2 compares diagnostic accuracies in individual study phases. Simple comparisons show that the presence of diagnostic proposals significantly increased the diagnostic accuracy (phase B compared to A – $p<0.001$, phase C compared to A – $p<0.001$). Nevertheless, in the case of a single proposal (either correct or incorrect) the accuracy was increased in interpretations with correct diagnostic

proposals, while substantially reduced with incorrect diagnostic proposals. Agreement rate with correct diagnostic proposals was very high in both CF and non-CF (89.6% and 87.5% respectively). With incorrect diagnostic proposals, the agreement rate reached 30.7% in CFs and even higher at 42.3% in non-CFs (although the difference between CFs and non-CFs did not reach statistical significance, p=0.13) (Figure 3). Hence interestingly, while CFs outperformed non-CFs regardless of the presence or absence of diagnostic proposals, there was no difference in diagnostic accuracy between the groups when correct diagnostic proposals were presented (*p=0.801*). Whilst both groups performed best when presented with correct diagnostic proposals, both groups performed second-best when presented with multiple diagnostic proposals (when one of those proposals is correct).

3.2. Presence and Absence of Diagnostic Proposals

Figure 4 shows differences in the fellows' performance when they were presented with diagnostic proposals compared to their baseline performance when no diagnostic proposals were presented. The performance by non-CFs exhibits more pronounced change compared to CFs when diagnostic proposals were provided. Incorrect diagnostic proposals had a negative impact on the performance of both CFs and non-CFs.

3.3. Paired Testing

A series of paired tests were performed to compare each approach within each cohort. Each approach (i.e. ECGs presented with and without diagnostic proposals) had a statistically significant ($p<0.05$) impact on the interpreter's diagnostic accuracy. This was true for both CFs and non-CFs.

3.4. Confidence ratings

Figure 5 shows that CFs had consistently greater confidence when correctly interpreting the ECG in comparison to non-CFs ($p<0.001$). This was also true regardless of whether correct, incorrect or multiple diagnostic proposals were presented. Interestingly, in comparison to non-

CFs, CFs remain more confident even when they interpreted the ECGs in a 'dangerously incorrect' way.

The confidence ratings were higher when CFs correctly interpreted ECGs for which one diagnostic proposal (either correct or incorrect) was offered (mean confidence ratings: without-diagnostic-proposals=7.95±1.87 vs. with-diagnostic-proposals=8.17±1.90, $p<0.001$). The same was also true for non-CFs (mean confidence ratings: without-diagnostic-proposals=6.96±2.14 vs. with-diagnostic-proposals=7.36±2.12, $p<0.001$).

In both groups, the confidence ratings were lower when an incorrect diagnostic proposal was provided (even when the interpreter was correct). There was a statistically significant (but subtle) difference in the confidence levels when CFs correctly interpreted ECGs that offered correct diagnostic proposals compared to when they correctly interpreted ECGs that offered incorrect proposals (mean confidence ratings: with-correct-diagnostic-proposals=8.40±1.78 vs. with-incorrect-diagnostic-proposals=7.76±4.04, $p<0.001$). The same was true for non-CFs (mean confidence ratings: with-correct-diagnostic-proposals=7.47±2.13 vs. with-incorrect-diagnostic-proposals=7.05±2.04, $p<0.001$).

Self-rated confidence poorly correlated with interpretation performance ($rho\approx2$, $p<0.001$). (Spearman correlation coefficients are shown in Table 3). The most significant correlation was observed when CFs interpreted ECGs with multiple diagnostic proposals ($rho=0.23$, $p<0.001$). The correlation was even stronger for diagnostic accuracies correlated to mean confidence ratings of CFs ($rho=0.30$, $p=0.28$) for only ECG interpretations with multiple diagnostic proposals. A similar result was found for non-CFs ($rho=0.33$, $p=0.23$).

3.5. Odds Ratios

Table 4 presents the odds ratios (ORs) for each independent/exposure variable as determined by the logistic regression model. Not surprisingly, the model found that being a non-CF did reduce the likelihood (OR=0.74) of correct interpretation. There was only a slightly greater likelihood of the interpreter being correct per month increase in the interpreter's experience (OR=1.01) and per unit increase in the confidence level (OR=1.19). Also not surprisingly, the interpreters were most likely correct when the ECG was presented with a correct diagnostic proposal (OR=10.87). However, somewhat unexpectedly, there was also a significant increase

(with the second highest OR) in the likelihood that the interpreter is correct when multiple diagnostic proposals were presented (OR=4.43).

4. DISCUSSION

The study leads to both expected and unexpected conclusions. Overall, the diagnostic accuracy was higher in the presence of diagnostic proposals. Nevertheless, a more detailed assessment showed that while the accuracy increased when correct proposals were presented, it was substantially reduced when ECGs were presented with incorrect diagnostic proposals (Figure 2). Analysis of the agreement rate with the diagnostic proposals confirmed that diagnostic proposals were very often accepted regardless of whether they were correct or incorrect. This trend was much more pronounced in non-CFs compared to CFs (Figure 3). Comparison with interpretations without any diagnostic proposals confirmed that the presence of incorrect proposals significantly reduced the interpreters' performance (Figure 4).

Previously, much smaller studies assessed the effect of incorrect computer-based ECG interpretations on the clinical decision making of the physician. For example, a group of 30 residents interpreting 23 ECG tracings with or without computerized diagnoses was significantly influenced by incorrect advices.[4] In addition. an erroneous computer interpretation of one ECG tracing accompanied with short clinical presentation assessed by 110 residents affected the aggressiveness of the prescribed treatment.[5] Our results confirm previous hypotheses and results by emphasizing the substantial influence a single diagnostic proposal can have, i.e. the fact that it can dangerously reduce the diagnostic accuracy of human interpretation when the computerized diagnosis is incorrect.

Interestingly, the highest accuracy was achieved by both groups in the third phase when ECGs were accompanied with multiple diagnostic proposals. Multivariate analysis showed that there was a significant increase in the likelihood that the interpreter is correct when multiple diagnostic proposals were presented (OR=4.43). Of course, some influence and "self-training" during the preceding two phases cannot be fully excluded. Nevertheless, there were substantial time gaps between the phases and the participants received no feedback until the study completion.

Our study also analyzed the self-rated confidence of ECG interpretations. The confidence ratings were higher when only one diagnostic proposal was present while the self-rated confidence poorly correlated with the interpretation accuracy. On the contrary, the most

significant correlation was found in the experiment with ECGs offering multiple proposals. While the overall confidence was somewhat lower in this phase, the increased correlation of the confidence with the interpretation accuracy suggested that the provision of multiple diagnostic choices made both the CFs and non-CFs subconsciously more diligent and impartial in the judgment of the ECG tracings. Therefore, this study indicates that different modes of presenting computerized diagnostics have an influence on clinical decision making. Consequently, these modes might potentially have a detrimental impact on the patient clinical pathway and outcome. Decision making researchers have previously described that external suggestions induce cognitive biases such as anchoring and confirmation bias which have a potent sub-conscious influence on the decision maker.[7,8] Our study supports this theory and confirms that an interpreter is relatively easily influenced and anchored by a single computerized diagnosis. We also note that the provision of only one diagnostic proposal is the current approach by the vast majority of manufacturers of electrocardiographic equipment.

Perhaps more interestingly, the interpreters had relatively good diagnostic accuracy when they were provided with multiple diagnostic proposals. This is likely to the fact that the provision of multiple suggestions removes the anchoring bias (specifically the propensity of humans to be biased towards readily available information or suggested conclusions). The provision of multiple diagnostic proposals encourages “System 2 thinking” (conscious deliberate reasoning which incites a less biased differential diagnosis) as opposed to “System 1 thinking” (automatic intuition [or ‘knee-jerk reactions’]).[7] Previous eye tracking studies have shown that experts are prone to use System 1 thinking when interpreting ECGs.[9]

In summary, our study might be interpreted as a recommendation to present multiple computerized diagnoses with each ECG tracing since this acts as a cognitive 'de-biasing strategy'. An extension to this de-biasing strategy might perhaps involve the use of interactive response technology where numerous independent decisions can be made from multiple options and automatically aggregated during a live session. However, this requires an unrealistic amount of time and resources especially at the point of care. [10]

4.1. Limitations

Several limitations of our study need to be recognized. The study was conducted among junior fellows with limited ECG interpretation experience. While it would be informative to conduct a similar study among cardiology and internal medicine consultants, this is not possible because of human resources reasons. The subgroups of readers were not matched

using any parameter; the years of training of the CFs were slightly longer when compared to NCFs (however multivariate logistic regression analysis did not show that this was confounding). No other characteristics of the medical competence of the fellows were used. Moreover, ECG interpretations were requested without any contextual information, i.e. patient history and/or other clinical data. Multiple diagnostic approaches in Phase 3 always involved one correct proposal which might not be fully realistic with fully automatic systems. The single diagnostic proposal (Phase 2) were correct and incorrect in exactly 50% of cases which does not necessarily reflect the diagnostic accuracy of diagnostic algorithms in presently available equipment (in reality, the proportions of correct and incorrect computerized diagnoses also depend on the abnormalities of the diagnosed tracings). The numbers of ECG tracings in categories shown in Table 1 were too small for any meaningful sub-analysis.

Authorship

TN - the conception and design of the study, interpretation of data, drafting and revising the article; RB - the conception and design of the study, analysis and interpretation of data, revising the article; IA, MS - the conception and design of the study, interpretation of data; LK - acquisition of data; DF, DG – design of the study, revising the article; JS – final approval; MM - the conception and design of the study, interpretation and analysis of data, revising the article.

Author Contribution

Tomas Novotny - the conception and design of the study, interpretation of data, drafting and revising the article; Raymond Bond - the conception and design of the study, analysis and interpretation of data, revising the article; Irena Andrsova, Martina Sisakova - the conception and design of the study, interpretation of data; Lumir Koc - acquisition of data; Dewar Finlay, Daniel Guldenring – design of the study, revising the article; Jindrich Spinar – final approval; Marek Malik - the conception and design of the study, interpretation and analysis of data, revising the article.

Conclitct of interest: none

Acknowledgments

The study was supported in part by the project (Ministry of Health, Czech Republic) for conceptual development of research organization 65269705 (University Hospital Brno, Brno, Czech Republic).

Summary table

What was already known on this topic:

- contemporary 12-lead electrocardiogram (ECG) devices offer automatic computerized diagnostic proposals to assist diagnostic decision-making
- the reliability of these automated diagnostic proposals is still sub-optimal.
- several smaller studies have suggested that incorrect computer advice can influence physician decision-making and can lead to additional unnecessary diagnostic testing and/or inappropriate therapy

What this study added to our knowledge:

- based on a total of 9000 ECG interpretations it was shown that computerized diagnostic proposals affect both the diagnostic accuracy and the interpreters´ confidence in their conclusive diagnosis
- the accuracy is significantly influenced especially when a single diagnostic proposal (either correct or incorrect) is provided - giving the correct diagnosis improves the accuracy while giving a wrong diagnosis lowers the accuracy
- presentation of multiple computerized diagnoses improved the diagnostic accuracy of ECG interpreters
- perhaps the presentation of multiple diagnostic choices together with the presentation of an algorithmic likelihood score should be considered in future models of automated diagnostic statements provided by electrocardiographic devices

References

1. Estes NA 3rd. Computerized interpretation of ECGs. Supplement not a substitute. Circ Arrhythm Electrophysiol 2013;6:2-4.

2. Willems JL, Abreu-Lima C, Arnaud P, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. N Engl J Med 1991;325:1767–73.

3. Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. J Electrocardiol 2007;40:385–90.

4. Hakacova N, Trägårdh-Johansson E, Wagner GS, Maynard C, Pahlm O. Computer-based rhythm diagnosis and its possible influence on nonexpert electrocardiogram readers. J Electrocardiol 2012;45:18–22.

5. Kligfield P, Gettes LS, Bailey JJ, et al. American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; American College of Cardiology Foundation; Heart Rhythm Society. Recommendations for the standardization and interpretation of the electrocardiogram: part I: The electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. Circulation 2007;115:1306–24.

6. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. J Am Med Inform Assoc 2003;10:478-83.

7. Southern WN, Arnsten JH. The effect of erroneous computer interpretation of ECGs on resident decision making. Med Decis Making 2009;29:372-6.

8. Novotny T, Bond RR, Andrsova I, et al. Data analysis of diagnostic accuracies in 12-lead electrocardiogram interpretation by junior medical fellows. J Electrocardiol 2015;48:988-94.

9. Stanovich KE, West, RF. Individual differences in reasoning: Implications for the rationality debate. Behavioral & Brain Sciences, 2000;23:645-665.

10. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science 1974;185:1124-31.

11. Bond RR, Zhu T, Finlay D, et al. Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: an objective quantitative approach. J Electrocardiol 2014;47:895-906.

12. Peace A, Ramsewak A, Cairns A, et al. Using computerised interactive response technology to assess electrocardiographers and for aggregating diagnoses. J Electrocardiol 2015;48:995-9.

Figure legends

Figure 1. Examples of ECGs used in the study (Panels A and B show non-life threatening conditions; panels C a life-threatening condition). The interpretation examples are coded as (1) correct, (2) almost correct, (3) incorrect, and (4) dangerously incorrect. Panel **A**: (1) ventricular preexcitation, (2) Wolf – Parkinson – White (WPW) syndrome, (3) non-specific intraventricular conduction disorder, (4) sub-acute myocardial infarction with ST elevations, inferior wall. Panel **B**: (1) atrial-triggered ventricular pacing, (2) ventricular pacing, (3) atrial pacing, (4) pacemaker dysfunction. Panel **C**: (1) acute myocardial infarction with ST elevations, lateral wall, right bundle branch block, (2) acute myocardial infarction with ST elevations, lateral wall, (3) acute myocardial infarction without ST elevations, anterior wall, (4) right bundle branch block.

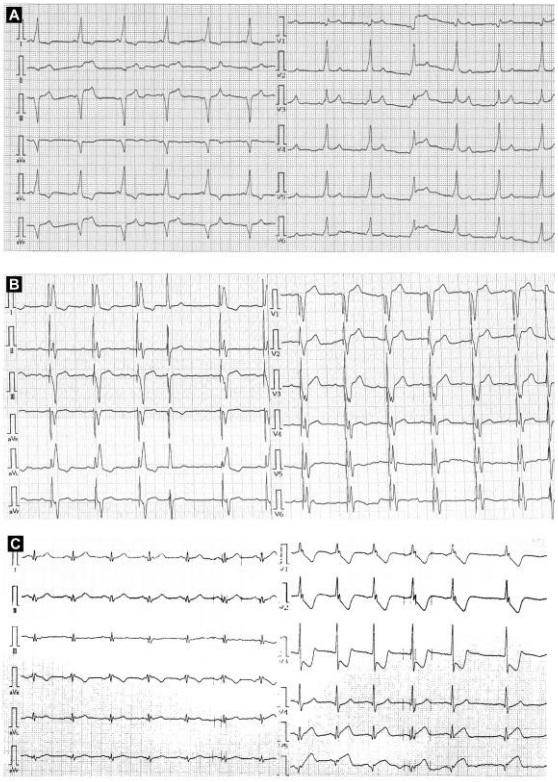

Figure 2. Percentage rates of ECG interpretations that were classified as (A) correct, (B) almost correct, (C) incorrect and (D) dangerously incorrect for Cardiology Fellows and Non-Cardiology Fellows when presented without diagnostic proposals, with one diagnostic

proposal (either correct or incorrect), with multiple diagnostic proposals, with correct diagnostic proposals, and with incorrect diagnostic proposals. The boxes represent interquartile ranges (IQRs); the central lines represent the medians, and the whiskers represent the minimum and maximum values (unless these values were greater than 1.5*IQR). Open circles show outliers outside the 1.5*IQR interval.

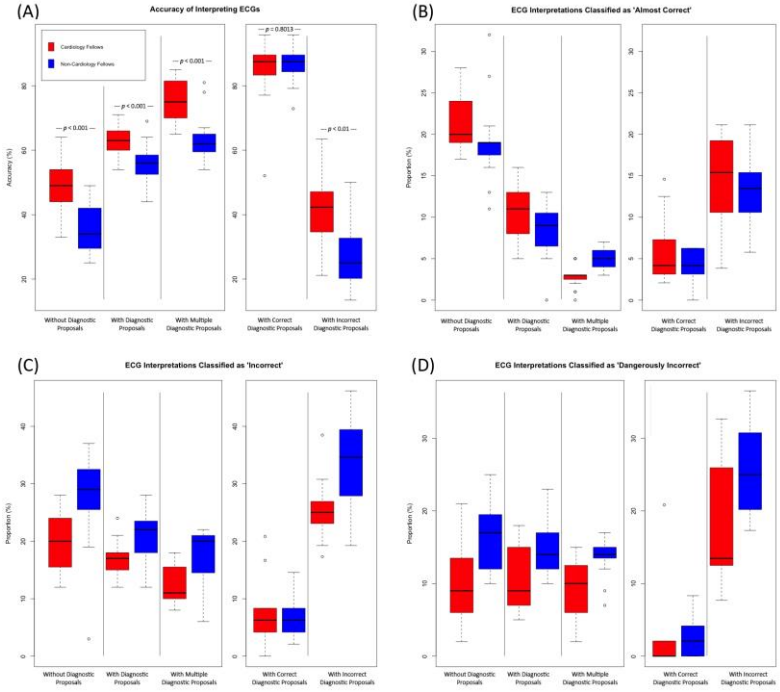

Figure 3. Agreement rate with diagnostic proposals when interpreting ECGs with one diagnostic proposal (either correct or incorrect). See Figure 2 for layout explanation.

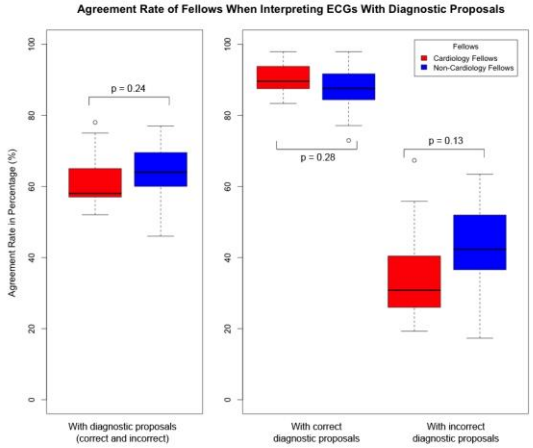

Figure 4. The differences between individual performance at baseline (i.e. interpretation performance when no diagnostic proposal is presented) and individual performance with one diagnostic proposal (either correct or incorrect), with multiple diagnostic proposals, with

correct diagnostic proposals and with incorrect diagnostic proposals. See Figure 2 for layout explanation.

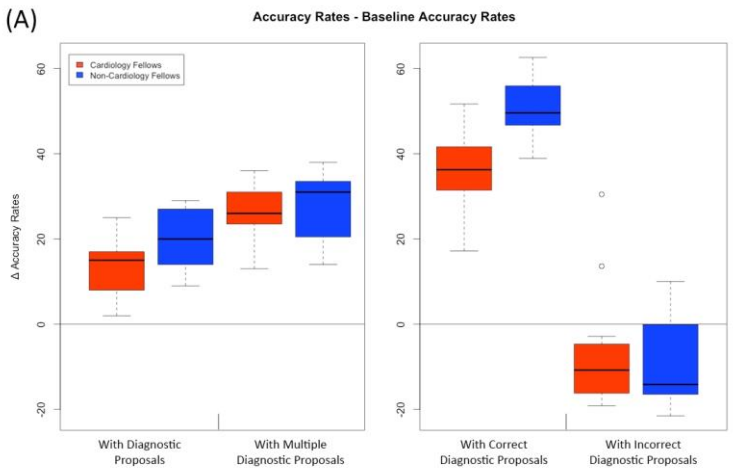

Figure 5. Self-rated confidence levels of Cardiology Fellows and Non Cardiology Fellows when their ECG interpretation was (A) correct, (B) almost correct, (C) incorrect and (D) dangerously incorrect and when interpreting ECG tracings without diagnostic proposals, with one diagnostic proposal, with multiple diagnostic proposals, with correct diagnostic proposals and with incorrect diagnostic proposals. See Figure 2 for layout explanation.

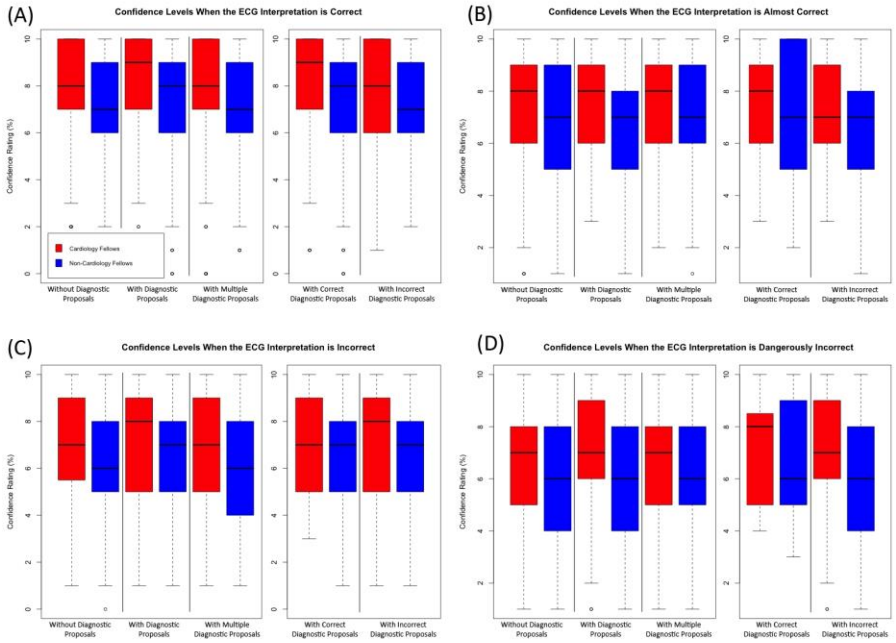

Table 1. Diagnoses of the selected 12-lead ECGs

ID	Diagnostic Class	# ECGs
1	Acute coronary syndrome	10
2	VT/IVR/WCT	5
3	Asystole, SA arrest	2
4	Other (non-acute) CAD	7
5	AVB	11
6	Intraventricular conduction disorder	10
7	APC/VPC	8
8	AF/AFl	12
9	SVT	2
10	Other	14
11	Paced rhythm	8
12	Normal	11
	Total:	**100**

VT – ventricular tachycardia, IVR – idioventricular rhythm, WCT – wide complex tachycardia, SA arrest – sinoatrial arrest, CAD – coronary artery disease, AVB – atrioventricular blockade, APC – atrial premature complex, VPC – ventricular premature complex, AF – atrial fibrillation, AFl – atrial flutter, SVT – supraventricular tachycardia, Other – long QT syndrome, accelerated junctional rhythm, atrial ectopic rhythm, left ventricular hypertrophy, pericarditis, preexcitation, digitalis toxicity, hyperkalemia, P

mitrale.

Table 2. Participants of the study

	#	Age	Gender	Months of Experience
Cardiology Fellows	15	30±2	(3 M, 12 F)	36±11
Non-Cardiology Fellows	15	28±2	(5 M, 10 F)	28±13
All Fellows	**30**	**29± 2**	**(8 M, 22 F)**	**32±12**

Non-cardiology Fellows included fellows of haematology, oncology, general internal medicine, and gastroenterology

Table 3. Spearman correlation between interpretation scores and confidence levels.

	Cardiology Fellows	Non-Cardiology Fellows
Without diagnostic proposals	*rho* = 0.21, $p<0.001$	*rho* = 0.21, $p<0.001$
With correct diagnostic proposals	*rho* = 0.20, $p<0.001$	*rho* = 0.12, $p<0.01$
With incorrect diagnostic proposals	*rho* = 0.12, $p<0.001$	*rho* = 0.15, $p<0.001$
With multiple diagnostic proposals	*rho* = 0.23, $p<0.001$	*rho* = 0.18, $p<0.001$

Correlations are between interpretation scores and confidence levels, however, the score was reversed to provide positive correlation coefficients (where 0=dangerously incorrect, 1=incorrect, 2=almost correct and 3=correct).

Table 4. Odds Ratios (ORs) per unit increase in each of the independent (exposure) variables.

Exposure Variable	Odds Ratio	95% CI	Std. Error	Z-value	P-value
Designation (NCFs)	0.74	0.67, 0.82	0.05	-5.71	< 0.001
Experience (Months)	1.01	1.00, 1.01	0.002	2.71	< 0.01
Without diagnostic proposals	1.40	1.23, 1.60	0.06	5.11	< 0.001
Correct diagnostic proposals	10.87	9.06, 13.09	0.09	25.43	< 0.001
Multiple diagnostic proposals	4.43	3.88, 5.06	0.06	21.86	< 0.001
Confidence-rating	1.19	1.16, 1.22	0.01	15.22	< 0.001